

Solutions: Examining Relationships

Checkpoint 2

Question 1

High blood pressure is unhealthy. Here are the results of one of the studies that link high blood pressure to death from cardiovascular disease. The researchers classified a group of white males aged 35 to 64 as having Low blood pressure or High, then followed the subjects for five years. The following two-way table gives the results of the study:

Cardiovascular death?	Blood pressure		Total
	Low	High	
Yes	21	55	76
No	2655	3283	5938
Total	2676	3338	6014

In this example, it would be appropriate to calculate:

- ☐ (a) conditional row percentages
- ☐ (b) conditional column percentages
- ☐ (c) the correlation coefficient r
- ☐ (d) the five-number summary of both variables
- ☐ (e) none of the above

Correct answer: (b)

Question 2

A local ice cream shop kept track of the number of cans of cold soda it sold each day, and the temperature that day, for two months during the summer. The data are displayed in the scatterplot below:

The one outlier corresponds to a day on which the refrigerator for the soda was broken. Which of the following is true?

- ☐ (a) A reasonable value of the correlation coefficient r for these data is 1.2.
- ☐ (b) If the temperature were measured in degrees Celsius ($C = 5/9(F-32)$), the value of r would change accordingly.
- ☐ (c) If the outlier were removed, r would increase.
- ☐ (d) If the outlier were removed, r would decrease.
- ☐ (e) Both (b) and (c) are correct.

Correct answer: (c)

Question 3

Suppose that the correlation r between two quantitative variables was found to be $r = 0$. This means that:

- ☐ (a) there is a strong linear relationship between the two variables.
- ☐ (b) there is no linear relationship between the two variables.
- ☐ (c) there is a strong relationship between the two variables.
- ☐ (d) there is no relationship between the two variables.
- ☐ (e) none of the above.

Correct answer: (b)

Question 4

A correlation of $r = .85$ is found between weekly sales of firewood and cough drops over a 1-year period. Which of the following is true?

- ☐ (a) There is a pretty strong positive linear relationship between sales of firewood and cough drops.
- ☐ (b) Fire must be the cause of coughing.
- ☐ (c) Temperature is a possible lurking variable that is “behind” this relationship.
- ☐ (d) Both (a) and (c) are true.
- ☐ (e) None of the above.

Correct answer: (c)

Question 5

The data in the scatterplot below are an individual's weight and the time it takes (in seconds) on a treadmill to raise his or her pulse rate to 140 beats per minute. The o's correspond to females and the +'s to males. Which of the following conclusions is most accurate?

- ☐ (a) There is a positive correlation between gender and weight, since men tend to weigh more.
- ☐ (b) There is a negative correlation between time and weight for males and for females.
- ☐ (c) There is a positive correlation between time and weight for males and for females.
- ☐ (d) Both (a) and (b) are correct.
- ☐ (e) Both (a) and (c) are correct.

Correct answer: (b)

Question 6

What can we say about the relationship between the correlation r and the slope b of the least-squares line for the same set of data?

- ☐ (a) Both r and b always have values between -1 and 1.
- ☐ (b) r is always larger than b .
- ☐ (c) r and b have the same sign (+ or -).
- ☐ (d) the slope b is always equal to the square of the correlation r .
- ☐ (e) b is always larger than r .

Correct answer: (c)

Question 7

A study was done on the timeliness of flights (on-time vs. delayed) of two major airlines: StatsAir and AirMedian. Data were collected over a period of time from five major cities and it was found that StatsAir does better overall (i.e., has a smaller percentage of delayed flights). However, in each of the five cities separately, AirMedian does better.

Which of the following is correct?

- ☐ (a) This situation is mathematically impossible.
- ☐ (b) This is an example of Simpson's Paradox.
- ☐ (c) "City" is a lurking variable in this example.
- ☐ (d) This is an example of a negative association between variables.
- ☐ (e) Both (b) and (c) are correct.

Correct answer: (e)

Please answer the question below. Your response will not be graded, but it will be available for your instructor to read.

Question 8

When faced with a data analysis problem that involves two variables, explain how you would decide which graphical display and numerical measures to use.

0 points